

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 February 2006 (23.02.2006)

PCT

(10) International Publication Number  
**WO 2006/019350 A1**

(51) International Patent Classification<sup>7</sup>: **G06K 9/00**,  
G06T 15/00

(21) International Application Number:  
PCT/SE2005/001203

(22) International Filing Date: 11 August 2005 (11.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0402048-3 19 August 2004 (19.08.2004) SE  
60/603,266 23 August 2004 (23.08.2004) US

(71) Applicant (for all designated States except US):  
**GROUND TRUTH VISION AB** [SE/SE]; Anckar-  
gripsgatan 3, S-211 19 Malmö (SE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SOLEM, Jan Erik**  
[SE/SE]; Andréégatan 9 A, S-211 49 Malmö (SE). **KAHL,**  
**Fredrik** [SE/SE]; Ostra Köpevägen 25, S-372 94 Listerby  
(SE).

(74) Agent: **VALEA AB**; Lindholmospiren 5, S-417 56 Göte-  
borg (SE).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,  
KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,  
MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ,  
OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL,  
SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,  
VN, YU, ZA, ZM, ZW.

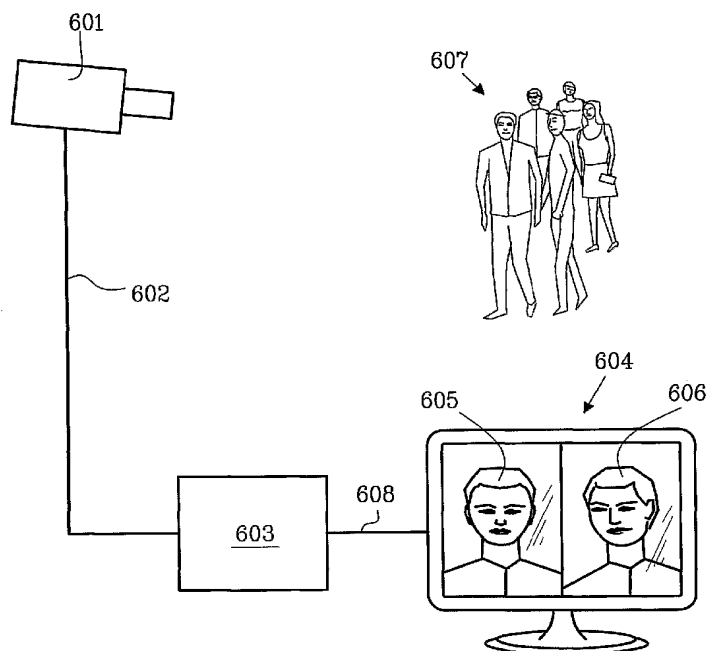
(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,  
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,  
RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,  
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with international search report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: 3D OBJECT RECOGNITION



(57) Abstract: A method, device, system, and computer program for object recognition of a 3D object of a certain object class using a statistical shape model for recovering 3D shapes from a 2D representation of the 3D object and comparing the recovered 3D shape with known 3D to 2D representations of at least one object of the object class.

WO 2006/019350 A1

### 3D object recognition

#### 5 ***Field of the invention***

The present invention relates to automated object recognition and in particular to automated object recognition of 3D objects using statistical shape information.

#### 10 ***Background of the invention***

There exist extremely reliable methods for personal identification using biometric data such as e.g. fingerprints, retinal patterns or similar unique features of the subject that rely on the cooperation of the subject. Face recognition may be an effective way of identifying a person without the cooperation or knowledge of the person. There are two main general problems for a face recognition system; *identifying* a person, i.e. determine the identity from images, and *verifying* the identity of a person, i.e. to certify that the person is who he/she claims to be. Specific applications are e.g. immigration, ID-cards, passports, computer logon, intranet security, video surveillance and access systems. The present invention aims at increasing the performance and efficiency of such systems using geometric information available through the use of statistical shape models.

In the area of statistical shape models, the invention is related to the Active Shape Models (ASM), introduced by Cootes and Taylor, ([1]: Cootes T.F. and Taylor C.J., *Active Shape Model Search using Local Grey-level Models: A Quantitative Evaluation*, British Machine Vision Conference, p.639-648, 1993). One distinction is that ASM have been used for inferring 2D shape from 2D observations or 3D shape from 3D observations whereas the invention uses 2D observations, i.e. images, to infer 3D shape. Also the observations are from multiple views (one or more imaging devices), something that is not handled in standard ASM. Cootes and Taylor have a number of patents in the area, the most relevant are (WO02103618A1 – Statistical Model) where parameterisation of 2D or 3D shapes are treated, (WO0135326A1 – Object Class Identification, Verification or Object Image Synthesis) where an object class is identified in images and (WO02097720A1 – Object Identification) in which objects are identified using modified versions of ASM and related techniques. Also related is Cootes et al. ([2]: Cootes T.F., Wheeler G.V, Walker K.N and Taylor C.J., *View-based Active Appearance Models*, Image and Vision Computing, 20(9-10),

P17019PC-a1e.doc

2

p.657-664, 2002.) where multi-view models are used but no explicit or consistent 3D data is contained in the model. There are also methods for deforming a 3D model of the object to fit the 2D projections of the object in the images such as in Blanz and Vetter ([3]: Blanz V. and Vetter T., *Face Recognition Based on Fitting a 3D Morphable Model*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 25(9), p.1063-1073, 2003.). These methods are very computationally expensive and often require manual intervention. Related patents are US6556196/EP1039417 (Method and apparatus for the processing of images) which describes a method for morphing a 3D model so that it will be a 3D representation of the object in the image by minimizing the projection error in the image.

One common problem for image based recognition is detecting the 2D shape of the object in the image, i.e. finding the relevant image region. Recent methods for detecting objects in images usually involve scanning the whole image at different scales for object specific image patterns and then using a classifier to decide if the region is relevant or not. The latest developments suggest the use of Support Vector Machines (SVM) for this task. A key element is the extraction of image features, i.e. parts of the image such as corners, edges and other interest points. This is usually done using correlation based schemes using templates or edge based methods using image gradients. For an overview of methods for face detection and feature extraction, cf. Zhao and Chellappa ([4]: Zhao W., Chellappa R., Rosenfeld A and Phillips P.J, *Face Recognition: A Literature Survey*, Technical report CAR-TR-948, 2000.) and the references therein. In [4] a review of current image based methods for face recognition is also presented.

When using image based methods for identification and verification there are two major problems, *illumination variation* and *pose variation*. Illumination variation will affect all correlation based methods where parts of images are compared since the pixel values vary with changing illumination. Also specular reflections can give rise to high changes in pixel intensity. Pose variation occurs since the projection in the image can change dramatically as the object rotates. These two problems have been documented in many face recognition systems and are unavoidable when the images are acquired in uncontrolled environments. Most of the known methods fail to handle these problems robustly.

The illumination problem is handled by the invention since no image correlation or comparison of image parts is performed. Instead features such as corners which

P17019PC-a1e.doc

3

are robust to intensity changes are computed, which make the shape reconstruction, to a large extent, insensitive to illumination and specular reflections. The invention handles the pose problem by using any number of images with different pose for training the statistical model. Any subset of the images, as  
5 few as a single image, can then be used to infer the 3D shape of the object.

***Summary of the invention***

The invention consists of a statistical model of the shape variations in a class of  
10 objects relating the two-dimensional (2D) projection in images to the three-dimensional (3D) shape of the object and the use of the 3D shape information for identification or verification of the object. Furthermore, the present invention relates to an image processing device or system for implementing such a method. The process is fully automatic and may be used e.g. for biometric identification  
15 from face images or identification of objects in for instance airport security X-ray images. The recovered 3D shape is the most probable shape consistent with the 2D projections, i.e. the images. The statistical model needs a bank of data, denoted training data, where the 3D positions of the image features are known, in order to learn the parameters of the model. Such data sampling can be done using e.g.  
20 binocular or multi-view stereo or range scanners. Once the model parameters are learned, the 3D shape can be computed using one or several images. The 3D shape is then used, by means of the presented invention together with the 2D image data, to identify or verify the object as a particular instance of the object class, e.g. the face belonging to a certain individual. A positive (or negative) identification  
25 initiate proper action by means of the presented innovation.

In a preferred embodiment of the invention, a method for object recognition of a three dimensional (3D) object is presented, the method comprising the steps of:

- obtaining at least one two dimensional (2D) representation of the object;
- 30 – detecting image features in the obtained 2D representation;
- recovering a highly probable 3D shape of the object of a certain object class consistent with 2D images of the object using at least one obtained image where 2D features are detected and using a learned statistical multi-view shape model of the shape variation; and
- 35 – comparing the recovered 3D shape with a reference representation of at least one object of the object class.

P17019PC-a1e.doc

4

In the method, the recovered 3D shape may be a complete surface model.

Still in the method, the complete surface model may be inferred from 2D or 3D features.

5

In another aspect of the method according to the present invention, the object class may contain non-rigid objects and the statistical shape model may be learned using 2D and 3D data specific for possible deformations of the objects in the non-rigid object class.

10

The method may further comprise the step of identifying an individual object of an object class or aiding in the identification of an individual object using the recovered 3D shape.

15 The method may yet further comprise the step of verifying the identity of an individual object of an object class or aiding in the verification of the identity of an individual object using the recovered 3D shape.

20 The method may further comprise the step of: fitting a surface to the recovered 3D shape using a learned statistical shape model for the surface of the object in order to regularize the surface shape in a manner specific for the object class.

25 In the method the object may be one or several of: a human face, a human body, inner organ(s) of a human body, blood vessel, animal, inner organs of an animal, a tumor, manufactured product(s) from an industrial process, a vehicle, an aircraft, a ship, military object(s).

In the method the reference representation may be stored in at least one of a non-volatile memory, database server, and personal identification card.

30

In another embodiment of the present invention, a device for object recognition of a three dimensional (3D) object is presented, comprising:

- means for obtaining at least one two dimensional (2D) representation of the object;
- 35 – means for detecting image features in the obtained 2D representation;
- means for recovering a highly probable 3D shape of the object of a certain object class consistent with 2D images of the object (607) using one or

P17019PC-a1e.doc

5

more images where 2D features are detected and using a learned statistical multi-view shape model of the shape variation; and

- means for comparing the recovered 3D shape with a reference representation of at least one object of the object class.

5

In the device the recovered 3D shape may be a complete surface model and the complete surface model may be inferred from 2D or 3D features.

10 In the device the object class may contain non-rigid objects and the statistical shape model may be learned using 2D and 3D data specific for possible deformations of the objects in the non-rigid object class.

15 The device may further comprise means for identifying an individual object of an object class or aiding in the identification of an individual object using the recovered 3D shape.

The device may still further comprise means for verifying the identity of an individual object of an object class or aiding in the verification of the identity of an individual object using the recovered 3D shape.

20

The device may further comprising means for: fitting a surface to the recovered 3D shape using a learned statistical shape model for the surface of the object in order to regularize the surface shape in a manner specific for the object class.

25 In the device the object may be one or several of: a human face, a human body, inner organ(s) of a human body, blood vessel, animal, inner organs of an animal, a tumor, manufactured product(s) from an industrial process, a vehicle, an aircraft, a ship, military object(s).

30 In the device the recovered 3D shapes of blood vessels or organs recovered from 2D projections, e.g. using X-ray imaging may be used for navigating steerable catheters or aiding physicians by displaying the recovered 3D shape.

35 The recovered 3D shapes of facial features may be used in the device to identify or to verify an identity of an individual in an access control system or security system, resulting in an acceptance or rejection of the individual.

The device may further comprise an interface for communicating with a personal identification card wherein the reference representation is stored.

Yet another embodiment of the present invention, a computer program stored in a computer readable storage medium and executed in a computational unit for object recognition of a three dimensional (3D) object is presented, comprising:

- an instruction set for obtaining at least one externally acquired two dimensional (2D) representation of the object;
- an instruction set for detecting image features in the obtained 2D representation;
- an instruction set for recovering a highly probable 3D shape of the object of a certain object class consistent with 2D images of the object using one or more images where 2D features are detected and using a learned statistical multi-view shape model of the shape variation; and
- an instruction set for comparing the recovered 3D shape with a reference representation of at least one object of the object class.

The computer program may further comprise an instruction set for identifying and/or verifying an individual object of an object class or aiding in the identification and/or verification of the individual object using the recovered 3D shape.

In another embodiment of the present invention, a system for object recognition of a three dimensional (3D) object is presented, comprising:

- means for obtaining at least one two dimensional (2D) representation of the object;
- means for detecting image features in the obtained 2D representation;
- means for recovering a highly probable 3D shape of the object of a certain object class consistent with 2D images of the object using one or more images where 2D features are detected and using a learned statistical multi-view shape model of the shape variation;
- means for comparing the recovered 3D shape with a reference representation of at least one object of the object class; and
- means for responding to a result from the means for comparison.

The system may further comprise means for identifying and/or verifying an individual object of an object class or aiding in the identification and/or verification of the individual object using the recovered 3D shape.

In the system the reference representation may be stored in at least one of a non-volatile memory, database server, and personal identification card.

5

***Brief description of the drawings***

In the following the invention will be described in a non-limiting way and in more detail with reference to exemplary embodiments illustrated in the enclosed  
10 drawings, in which:

Fig. 1 illustrates a two-step procedure for recovering 3D data from an input image.

Fig. 2 illustrates a process of surface fitting to a recovered 3D shape.

15

Fig. 3 is a schematic block diagram of a device according to the present invention.

Fig. 4 illustrates a schematic block diagram of the steps of a method according to the present invention.

20

Fig. 5 is a schematic illustration of a system according to the present invention.

***Detailed description of the invention***

25 The invention consists of an image processing system for automatic recovery of 3D shape from images of objects belonging to a certain class. This 3D reconstruction is done by establishing a statistical shape model, denoted the feature model, that relates the 2D image features, e.g. interest points or curves, to their corresponding 3D positions. Such a model is learned, i.e. the model parameters are estimated,  
30 from training data where the 2D-3D correspondence is known. This learning phase may be done using any appropriate system for obtaining such 2D-3D correspondence, including, but not limited to binocular or multi-view image acquisition systems, range scanners or similar setups. In this process the object of interest is measured and a reference model of the object is obtained which may be  
35 used in subsequent image analysis as will be described below.



Given an input image, the process of recovering the 3D shape is a two-step procedure. First the image features such as points, curves and contours are found in the images e.g. using techniques such as e.g. ASM [1] or gradient based methods or classifiers such as SVM. Then the 3D shape is inferred using the learned  
 5 feature model. This is illustrated in Figure 1. Fig 1a illustrates an image of a face to be analysed, Fig. 1b illustrates the detection of object features to be used in the analysis and shape information process, and Fig. 1c illustrates the inferred 3D shape to be used in the recognition process.

10 There is also the option of extending the 3D shape representation from curves and points to a full surface model by fitting a surface to the 3D data. This is illustrated in Figure 2, where Fig. 2a illustrates the inferred 3D shape, 2b illustrates a fitted surface to the 3D data, and Fig. 2c illustrates a 3D rendered surface model of the fitted surface.

### 15 ***The Feature Model***

Suppose we have a number of elements in a  $d$ -dimensional vector  $\mathbf{t}$ , for example, a collection of 3D points in some normalized coordinate system. The starting point for  
 20 the derivation of the model is that the elements in  $\mathbf{t}$  can be related to some latent vector  $\mathbf{u}$  of dimension  $q$  where the relationship is linear:

$$\mathbf{t} = W\mathbf{u} + \mu \quad (1)$$

25 where  $W$  is a matrix of size  $d \times q$  and  $\mu$  is a  $d$ -vector allowing for non-zero mean. Once the model parameters  $W$  and  $\mu$  have been learned from examples, they are kept fix. However, our measurements take place in the images, which usually is a non-linear function of the 3D features according to the projection model for the relevant imaging device.

30 Denote the projection function with  $f : R^d \rightarrow R^e$ , projecting all 3D features to 2D image features, for one or more images. Also, we need to change coordinate system of the 3D features to suit the actual projection function. Denote this mapping by  $T : R^d \rightarrow R^d$ . Typically,  $T$  is a similarity transformation of the world  
 35 coordinate system. Thus,  $f(T(\mathbf{t}))$  will project all normalised 3D data to all images. Finally, a noise model needs to be specified. We assume that the image measurements are independent and normally distributed, likewise, the latent

variables are assumed to be Gaussian with unit variance  $\mathbf{u} \sim N(0, I)$ . Thus, in summary:

$$\mathbf{t}_{2D} = f(T(\mathbf{t})) + \varepsilon = f(T(W\mathbf{u} + \mu)) + \varepsilon \quad (2)$$

5

where  $\varepsilon \sim N(0, \sigma^2 I)$  for some scalar  $\sigma$ . The model is related to PPCA, cf. Tipping and Bishop ([5]: Tipping M.E., Bishop C.M., *Probabilistic Principal Component Analysis*, Phil. Trans. Royal Soc. London B, 61(3), p.611-622, 1999.), but there are also differences due to the non-linearity of  $f(\cdot)$ . Before the model can be used, its parameters need to be estimated from training data. Given that it is a probabilistic model, this is best done with maximum likelihood (ML). Suppose we are given  $n$  examples  $\{\mathbf{t}_{2D,i}\}_{i=1}^n$ , the ML estimate for  $W$  and  $\mu$  is obtained by minimizing:

10

$$\sum_{i=1}^n \left( \frac{1}{\sigma^2} \|\mathbf{t}_{2D,i} - f(T_i(\mathbf{u}_i))\|^2 + \|\mathbf{u}_i\|^2 \right) \quad (3)$$

15

over all unknowns. The standard deviation  $\sigma$  is estimated a priori from the data. Once the model parameters  $W$  and  $\mu$  have been learned from examples, they are kept fix. In practice, to minimize (3) we alternatively optimize over  $(W, \mu)$  and  $\{\mathbf{u}_i\}_{i=1}^n$  using gradient descent. Initial estimates can be obtained by intersecting 3D structure from each set of images and then applying PPCA algorithms for the linear part. The normalization  $T_i(\cdot)$  is chosen such that each normalized 3D sample has zero mean and unit variance.

20

There are three different types of geometric features embedded in the model.

25

**Points:** A 3D point which is visible in  $m > 1$  images will be represented in the vector  $\mathbf{t}$  with its 3D coordinates (X,Y,Z). For points visible in only one image,  $m=1$ , no depth information is available, and such points are represented similarly to apparent contour points.

30

**Curves:** A curve will be represented in the model by a number of points along the curve. In the training of the model, it is important to parameterize each 3D curve such that each point on the curve approximately corresponds to the same point on the corresponding curve in the other examples.

35

**Apparent contours:** As for curves, we sample the apparent contours (in the images). However, there is no 3D information available for the apparent contours as they are view-dependent. A simple way is to treat points of the apparent contours as 3D points with a constant, approximate (but crude) depth estimate.

5

### ***Finding Image Features***

In the on-line event of a new input sample, we want to automatically find the latent variables  $\mathbf{u}$  and, in turn, compute estimates of the 3D features  $\mathbf{t}$ . The missing component in the model is the relationship between 2D image features and the underlying grey-level (or colour) values at these pixels. There are several ways of solving this, e.g. using an ASM (denoted the *grey-level model*) or detector based approaches.

### 15 *The Grey-level Model*

Again, we adopt a linear model (PPCA). Using the same notation as in (1), but now with the subscript  $gl$  for grey-level, the model can be written

$$20 \quad \mathbf{t}_{gl} = \mathbf{W}_{gl} \mathbf{u}_{gl} + \mu_{gl} + \varepsilon_{gl} \quad (4)$$

where  $\mathbf{t}_{gl}$  is a vector containing the grey-level values of all the 2D image features and  $\varepsilon_{gl}$  is Gaussian noise in the measurements. In the training phase, each data sample of grey-levels is normalized by subtracting the mean and scaling to unit variance. The ML-estimate of  $\mathbf{W}_{gl}$  and  $\mu_{gl}$  is computed with the EM-algorithm [5].

25

### *Detector-based Methods*

Image interest points and curves can be found by analyzing the image gradient using e.g. the Harris corner-detector. Also, specially designed filters can be used as detectors for image features. By designing the filters so that the response for certain local image structures are high, image features can be found using a 2D convolution.

### 35 *Classification Methods*

Using classifiers such as SVM, image regions can be classified as corresponding to a certain feature or not. By combining a series of such classifiers, one for each image feature (points, curves, contours etc.) and scanning the image at all appropriate scales the image features can be extracted. Examples can be e.g. an eye detector  
5 for facial images.

### *Deformable Models*

Using a deformable model such as the Active Contour Models, also called snakes, of  
10 a certain image feature is very common in the field of image segmentation. Usually the features are curves. The process is iterative and tries to optimize an energy function. An initial curve is deformed gradually to the best fit according to an energy function that may contain terms regulating the smoothness of the fit as well  
15 as other properties of the curve.

### ***Surface Fitting to the 3D Data***

Once the 3D data is recovered, a surface model can be fitted to the 3D structure. This might be desirable in case the two-step procedure above only produces a  
20 sparse set of features in 3D space such as e.g. points and space curves. Even if these cues are characteristic for a particular sample (or individual), it is often not enough to infer a complete surface model, and in particular, this is difficult in the regions where the features are sparse. Therefore, a 3D surface model consisting of the complete mean surface is introduced. This will serve as a domain-specific, i.e.  
25 specific for a certain class of objects, regularizer. This approach requires that there is dense 3D shape information available for some training examples in the training data of the object class obtained from e.g. laser scans or in the case of medical images from e.g. MRI or computer tomography. From these dense 3D shapes, a model can be built separate from the feature model above. This means that, given  
30 recovered 3D shape, in the form of points and curves, from the feature model, the best dense shape according to the recovered 3D shape can be computed. This dense shape information can be used to improve surface fitting.

To illustrate with an example, consider the case of the object class being faces. The  
35 model is then learned using e.g. points, curves and contours in images together with the true 3D shape corresponding to these features obtained from e.g. multi-view stereo techniques. A second model is then created and learned using e.g. laser

P17019PC-a1e.doc

12

scans of faces, giving a set of face surfaces. This second model can be used to find the most probable (or at least highly probable) mean face surface (according to the second model) corresponding to the features or the recovered 3D shape. A surface can then be fitted to the 3D shape with the additional condition that where there is  
5 no recovered 3D shape, the surface should resemble the most probable mean face surface.

As a second example, consider the case of the object class being a particular blood vessel, e.g. the aorta. The model is then learned using e.g. curves and contours in  
10 images together with the true 3D shape obtained as e.g. a 3D MRI image. From the true 3D shapes a second model is learned comprising of the surface of the aorta. Then the most probable (or highly probable) aorta surface can be recovered from the image features or from the 3D shape recovered by the primary shape model.

15 The method provides the most probable or an at least highly probable 3D shape, in many applications this is sufficient and the identification and/or verification process is not necessary for the final application.

We have now described the underlying method used for verification and/or  
20 identification purposes. Referring now to Fig. 3 a description of a device 400 implementing the preferred method according to the present invention will be given. Such a device 400 may be any appropriate type of computational device such as, but not limited to, a personal computer (PC), workstation, embedded computer, or stand alone device with a computational unit 401, such as a  
25 microprocessor, DSP (digital signal processor), FPGA (field programmable gate array), or ASIC (application specific integrated circuit). The device 400 has some input means 404 for obtaining images for analysis and final identification and/or verification. The input means 404 may be of any suitable communication interface depending on image type and include, but is not limited to, USB (universal serial  
30 bus), frame grabber, Ethernet, or Firewire. Image data is transferred to a computational unit 401 wherein software for execution of the above described method according to the present invention resides. The device 400 may further comprise some volatile or non-volatile memory 402 containing information related to a reference material for comparison and/or analysis purposes, e.g. known 2D-3D  
35 relationships of objects of interest. The device 400 may still further comprise communication means for communicating with other computational devices over e.g. a network protocol (such as Ethernet or similar protocols) and output means

P17019PC-a1e.doc

13

405 for outputting results to for instance a screen for convenient viewing or to a control device (not shown) for controlling an external process wherein the objects of interest are part of. Such processes may include, but is not limited to, industrial production processes where objects may be selected or deselected depending on the result from the identification and/or verification method according to the present invention, security processes again for selection or deselection purposes in for instance airport security systems for examination of the contents of suitcases, bags or other luggage equipment, or medical applications where the recovered 3D shape may be used e.g. for navigation of instruments or medical devices.

10

The method for object recognition according to the present invention may be illustrated using Fig. 4. The method may comprise the following steps:

1. Obtaining at least one image of an object to be identified and/or verified (501);
2. Detecting image features, such as curves, points, and apparent contours (502).
3. Analysing the obtained image and inferring 3D shape corresponding to the image features, using a statistical shape model (503);
4. Comparing the analysis with reference images previously obtained and comparing the 3D shape in a sparse or dense form with reference 3D shape previously obtained (504); and
5. Responding to an output from the comparison process (505).

25 In another embodiment of the present invention a system is used for obtaining images, analyzing, and responding to results from the identification and/or verification process, as may be seen in Fig. 5. Such a system may include at least one image acquisition device 601, a computational device 400, 603 as described above, and some type of responsive equipment such as e.g. the industrial process equipment or the security process equipment described above. At least one image acquisition device 601 is used for acquiring one or more images which are transferred 602 to a computational device 603 for analysis and verification and/or identification. The result from this process is transmitted to a control system or display system 604. In the case of a face detection system at least one image of a person is obtained, for instance the face of the person, and the image or images are transmitted to the computational device 603, using any suitable communication means 602 (wired or wireless), for analysis and comparison of the acquired image

P17019PC-a1e.doc

14

or images with data obtained from reference measurements for example with known 2D-3D relationships; however, comparison may be made between an inferred 3D shape with a stored 3D reference data or between a 2D surface model with a stored 2D reference surface model. The result may be made available through for instance a display unit 604 and may for illustrative purposes be displayed with both a reference image 605 and the obtained image 606 or images rendered from the recovered 3D shape as shown in Fig. 5. It should be appreciated by the person skilled in the art that the image acquisition system and/or display/control system may be incorporated with the computational device forming an integral unit and that the result may be displayed in any suitable manner and is not limited to the above described example. Instead of transferring the result to a display unit 604 it may be used in any suitable control process for controlling e.g. an alarm, an entrance system, control gate, toll gate, and so on.

Some of the benefits the present invention contributes to the technical field may be illustrated with the following list:

- Any number of images, even as few as a single image, may be used to automatically recover the 3D shape of an object in the object class.
- A statistical multi-view model that represents 2D *and* 3D data consistently.
- The process is automatic and computationally efficient.
- The process is robust to illumination and specular reflections which is a problem for 3D reconstruction methods based on image correlation or photo-consistency.
- Surfaces can be fitted to the 3D structure using domain specific regularizers learned from statistical shape models.

The flexibility of the present invention may be illustrated with the following list:

- The statistical shape model may be used for any class of objects and the projection of these objects in images.
- The approach may be used for any kind of imaging device (camera, X-ray, multi-spectral, thermal, etc.).
- The invention may be used with any number of imaging devices (one or more).
- The invention includes the possibility of combining many different techniques for establishing 2D to 3D correspondence (image acquisition systems, range

P17019PC-a1e.doc

15

scanners, stereo image systems, multi-view stereo image systems, X-ray etc.)

- The invention includes the possibility of using different techniques, such as ASM, gradient based methods or deformable models, for finding the image features.
- If the object class contains non-rigid objects, the invention includes the possibility to establish 2D to 3D models for different deformations of the object (e.g. different facial expressions).
- The invention includes the possibility of using a statistical shape model for surface fitting to the recovered 3D shape.

The reference representations of objects may be stored in several different locations and with different types of systems, such as, but not limited to, locally on some non-volatile memory in a device utilizing the object recognition according to the present invention, in a centralized server, e.g. a database server, or a personal identification card containing a reference representation of an object such as a person and this identification card may be used in for instance an access system. Communication between an object recognition system and a reference representation storage system may be utilized with different types of security levels and/or schemes, such as RADIUS, DIAMETER, SSL, SSH, or any other encrypted communication system as understood by the person skilled in the art.

Possible application areas for the above described invention range from object identification and verification in industrial processes, determining and/or identifying objects for security reasons, object recognition for military purposes, e.g. automatic determination of military vehicles, military ships, aircrafts, and so on, face recognition systems for many different applications, e.g. biometrics, information security, law enforcement, smart cards, access control and so on.

The above mentioned and described embodiments are only given as examples and should not be limiting to the present invention. Other solutions, uses, objectives, and functions within the scope of the invention as claimed in the below described patent claims should be apparent for the person skilled in the art.



**Claims**

1. A method for object recognition of a three dimensional (3D) object,  
comprising the steps of:
  - 5       – obtaining at least one two dimensional (2D) representation of said  
          object;
  - detecting image features in said obtained 2D representation;
  - recovering a highly probable 3D shape of said object of a certain object  
10       class consistent with 2D images of said object using at least one  
          obtained image where 2D features are detected and using a learned  
          statistical multi-view shape model of the shape variation; and
  - comparing said recovered 3D shape with a reference representation of at  
          least one object of said object class.
- 15   2. The method according to claim 1, wherein said recovered 3D shape is a  
      complete surface model.
3. The method according to claim 2, wherein said complete surface model is  
20       inferred from 2D or 3D features.
4. The method according to claims 1 to 3, wherein said object class contains  
      non-rigid objects and said statistical shape model is learned using 2D and  
      3D data specific for possible deformations of the objects in said non-rigid  
25       object class.
5. The method according to any of claims 1 - 4, further comprising the step of  
      identifying an individual object of an object class or aiding in the  
      identification of an individual object using said recovered 3D shape.
- 30   6. The method according to any of claims 1 - 4, further comprising the step of  
      verifying the identity of an individual object of an object class or aiding in  
      the verification of the identity of an individual object using said recovered 3D  
      shape.
- 35   7. The method according to claim 2 or 3, further comprising the step of: fitting  
      a surface to said recovered 3D shape using a learned statistical shape model

P17019PC-a1e.doc

17

for said surface of the object in order to regularize said surface shape in a manner specific for said object class.

- 5           8. The method according to claim 1 - 7, wherein said object may be one or several of: a human face, a human body, inner organ(s) of a human body, blood vessel, animal, inner organs of an animal, a tumor, manufactured product(s) from an industrial process, a vehicle, an aircraft, a ship, military object(s).
- 10          9. The method according to claim 1, wherein said reference representation is stored in at least one of a non-volatile memory, database server, and personal identification card.
- 15          10. A device (400) for object recognition of a three dimensional (3D) object, comprising:
- means for obtaining (404) at least one two dimensional (2D) representation of said object (607);
  - means for detecting (401) image features in said obtained 2D representation;
  - 20       – means for recovering (401) a highly probable 3D shape of said object of a certain object class consistent with 2D images of said object (607) using one or more images where 2D features are detected and using a learned statistical multi-view shape model of the shape variation; and
  - means for comparing (401) said recovered 3D shape with a reference
  - 25       representation of at least one object of said object class.
11. The device (400) according to claim 10, wherein said recovered 3D shape is a complete surface model.
- 30          12. The device (400) according to claim 11, wherein said complete surface model is inferred from 2D or 3D features.
- 35          13. The device (400) according to claims 10 - 12, wherein said object class contains non-rigid objects and said statistical shape model is learned using 2D and 3D data specific for possible deformations of the objects in said non-rigid object class.

P17019PC-a1e.doc

18

14. The device (400) according to any of claims 10 - 13 further comprising means for identifying an individual object of an object class or aiding in the identification of an individual object using said recovered 3D shape.
- 5      15. The device (400) according to any of claims 10 - 13, further comprising means for verifying the identity of an individual object of an object class or aiding in the verification of the identity of an individual object using said recovered 3D shape.
- 10      16. The device (400) according to claim 11 or 12, further comprising means for: fitting a surface to said recovered 3D shape using a learned statistical shape model for said surface of the object in order to regularize said surface shape in a manner specific for said object class.
- 15      17. The device (400) according to claim 10 - 16, wherein said object may be one or several of: a human face, a human body, inner organ(s) of a human body, blood vessel, animal, inner organs of an animal, a tumor, manufactured product(s) from an industrial process, a vehicle, an aircraft, a ship, military object(s).
- 20      18. The device (400) according to claim 17, wherein said recovered 3D shapes of blood vessels or organs recovered from 2D projections, e.g. using X-ray imaging, are used for navigating steerable catheters or aiding physicians by displaying said recovered 3D shape.
- 25      19. The device (400) according to claim 17, wherein said recovered 3D shapes of facial features are used to identify or to verify an identity of an individual in an access control system or security system, resulting in an acceptance or rejection of said individual.
- 30      20. The device (400) according to claim 10, further comprising an interface for communicating with a personal identification card wherein said reference representation is stored.
- 35      21. A computer program stored in a computer readable storage medium (402) and executed in a computational unit (401) for object recognition of a three dimensional (3D) object, comprising:

P17019PC-a1e.doc

19

- an instruction set for obtaining at least one externally acquired two dimensional (2D) representation of said object (607);
- an instruction set for detecting image features in said obtained 2D representation;
- 5      – an instruction set for recovering a highly probable 3D shape of said object of a certain object class consistent with 2D images of said object (607) using one or more images where 2D features are detected and using a learned statistical multi-view shape model of the shape variation; and
- 10     – an instruction set for comparing said recovered 3D shape with a reference representation of at least one object of said object class;

22. The computer program according to claim 21, further comprising an instruction set for identifying and/or verifying an individual object of an object class or aiding in said identification and/or verification of said individual object using said recovered 3D shape.

23. A system for object recognition of a three dimensional (3D) object, comprising:
- 20     – means for obtaining (601) at least one two dimensional (2D) representation of said object (607);
  - means for detecting (401) image features in said obtained 2D representation;
  - means for recovering (401, 603) a highly probable 3D shape of said object of a certain object class consistent with 2D images of said object (607) using one or more images where 2D features are detected and using a learned statistical multi-view shape model of the shape variation;
  - 25     – means for comparing (401) said recovered 3D shape with a reference representation of at least one object of said object class; and
  - 30     – means for responding (604) to a result from said means for comparison (401).

24. The system according to claim 23, further comprising means for identifying and/or verifying an individual object of an object class or aiding in said identification and/or verification of said individual object using said recovered 3D shape.

P17019PC-a1e.doc

20

25. The system according to claim 23, wherein said reference representation is stored in at least one of a non-volatile memory, database server, and personal identification card.

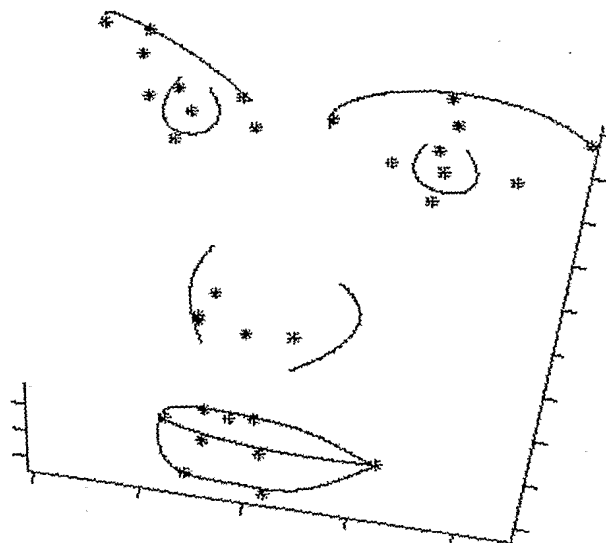
1/4



*Fig. 1a*

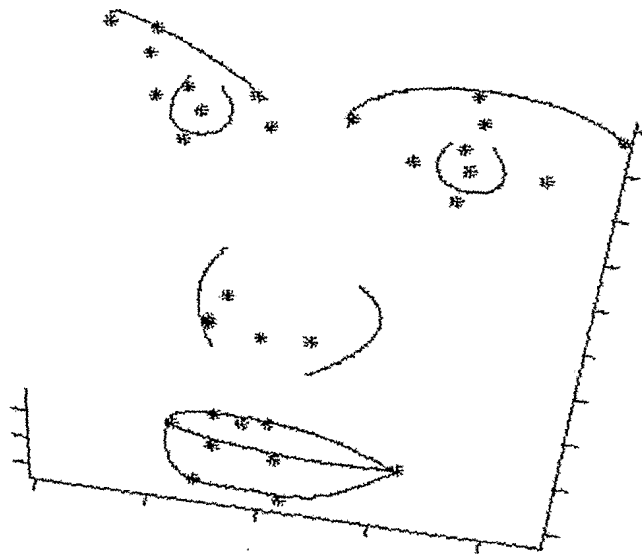


*Fig. 1b*

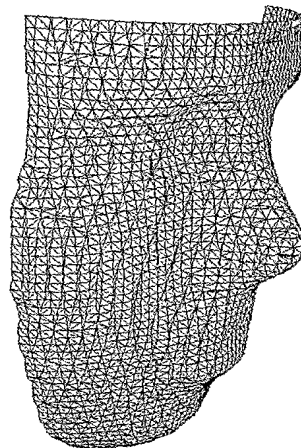


*Fig. 1c*

2/4



*Fig. 2a*

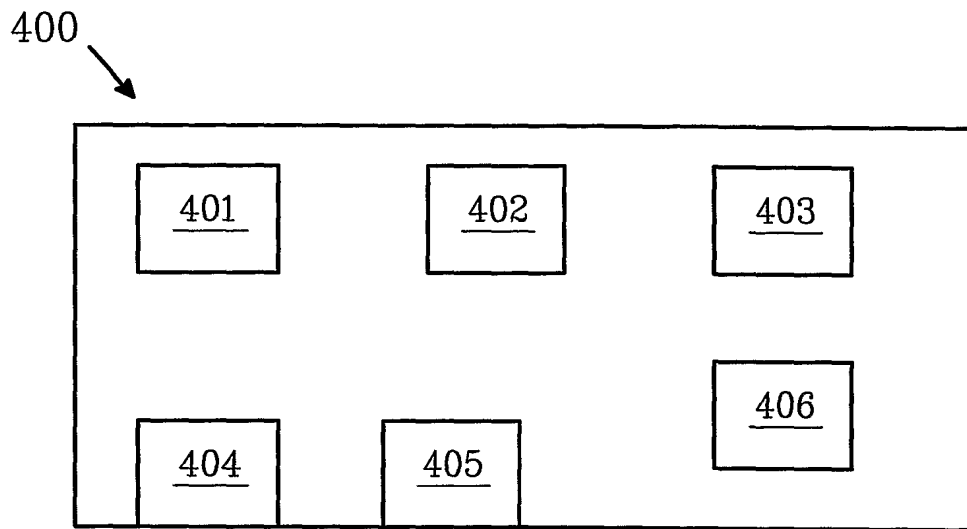
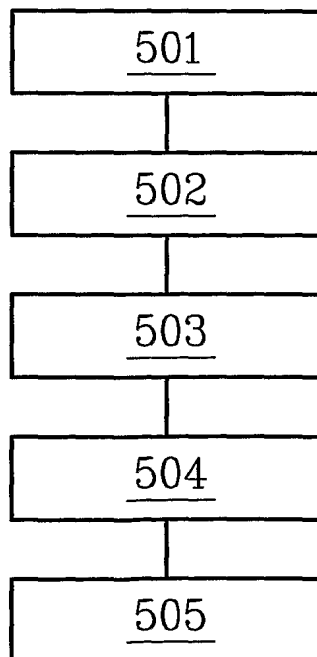


*Fig. 2b*

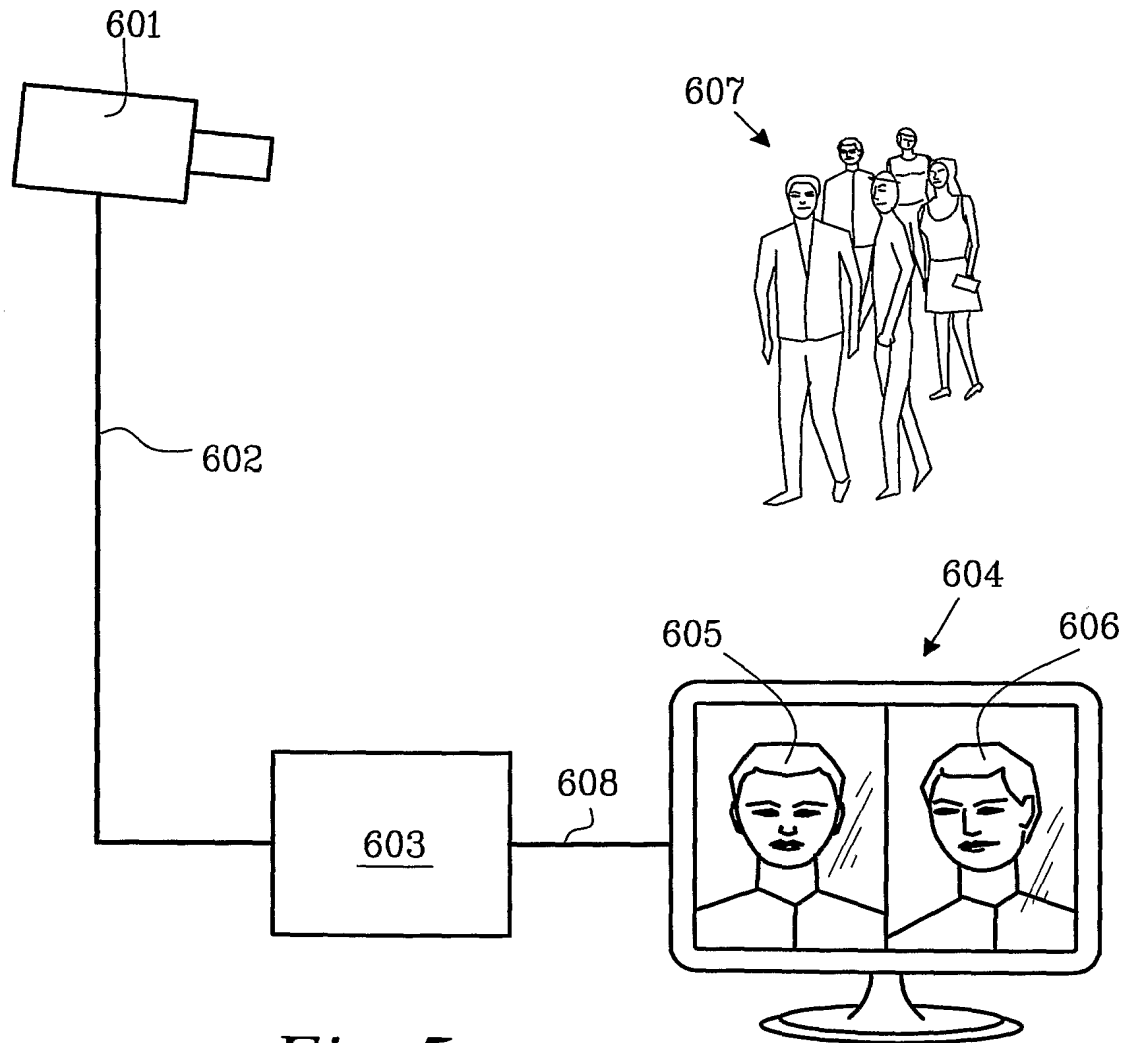


*Fig. 2c*

3/4

*Fig. 3**Fig. 4*



*Fig. 5*

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/SE 2005/001203

## A. CLASSIFICATION OF SUBJECT MATTER

IPC7: G06K 9/00, G06T 15/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: G06K, G06T, A61B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-INTERNAL, WPI DATA, PAJ

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 20010020946 A1 (YUICHI KAWAKAMI ET AL), 13 Sept 2001 (13.09.2001), figure 1, [0012]-[0021]; [0033]-[0047]	1-25
A	EP 1039417 A1 (MAX-PLANCK-GESELLSCHAFT), 27 Sept 2000 (27.09.2000), figures 1,5, [0008]-[0017];[0020]	1-25
A	US 6137896 A (SHOUDE CHANG ET AL), 24 October 2000 (24.10.2000), column 2, line 39 - column 3, line 41	1-25

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

\* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

7 November 2005

Date of mailing of the international search report

10-11-2005

Name and mailing address of the ISA/  
Swedish Patent Office  
Box 5055, S-102 42 STOCKHOLM  
Facsimile No. +46 8 666 02 86

Authorized officer

Ulrika Westman/MN  
Telephone No. +46 8 782 25 00

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

01/10/2005

International application No.

PCT/SE 2005/001203

US	20010020946	A1	13/09/2001	JP	2001256483	A	21/09/2001
EP	1039417	A1	27/09/2000	US	6556196	B	29/04/2003
US	6137896	A	24/10/2000	NONE			